Joint-MAP Tomographic Reconstruction with Patch Similarity Based Mixture Prior Model

Yang Chen, Yinsheng Li, Weimin Yu, Limin Luo, Wufan Chen, Christine Toumoulin

Joint-MAP Tomographic Reconstruction with Patch Similarity Based Mixture Prior Model

Yang Chen, Yinsheng Li, Weimin Yu, Limin Luo, Wufan Chen, Christine Toumoulin

Abstract

Tomographic reconstruction from noisy projections do not yield adequate results. Mathematically, this tomographic reconstruction represents an ill-imposed problem due to information missing caused from the presence of noise. Maximum A Posteriori (MAP) or Bayesian reconstruction methods offer possibilities to improve the image quality as compared with analytical methods in particular by introducing a prior to guide the reconstruction and regularize the noise. With an aim to achieve robust utilization of continuity/connectivity information and overcome the heuristic weight update for other nonlocal prior methods, this paper proposed a novel Patch Similarity based Mixture (PSM) prior model for tomographic reconstruction. This prior is defined by a weighted Gaussian distance between neighborhood intensities. The weight quantifies the similarity between local neighborhoods and is computed using a maximization entropy constraint. This prior is then introduced within a joint image/weight MAP CT reconstruction algorithm. Several acceleration trials including Compute Unified Device Architecture (CUDA) parallelization is applied to alleviate the intensive patch distance computation involved in the joint algorithm. The method was tested with both synthetic phantoms and clinical CT data and compared in accuracy with five other reconstruction algorithms which are FBP and Bayesian-based. Reconstruction results show that the proposed reconstructions are able to produce high-quality images with ensured iteration convergence.

Index Terms

Maximum a posteriori (MAP), Markov Random Fields (MRF), Patch Similarity Mixture (***PSM***) prior model, Tomographic Reconstruction, Compute Unified Device Architecture (CUDA).

I. INTRODUCTION

Statistical algorithms generally outperform conventional filtered back-projection (FBP) techniques due to the fact that they can provide a more accurate modelling of the data noise, system geometry and detector response [1-6].

Yang Chen, Yinsheng Li, Weimin Yu and Limin Luo are with the Laboratory of Image Science and Technology, Southeast University,Nanjing 210096, China

Christine Toumoulin is with INSERM U642, Laboratoire LTSI, Université de Rennes I, Rennes, F-35000, France

Yang Chen is also with INSERM U642, Laboratoire LTSI, Université de Rennes I, Rennes, F-35000, France

Yang Chen, Limin Luo and Christine Toumoulin are also with the Centre de Recherche en Information Biomedicale Sino-Francais (LIA CRIBs), Rennes, F-35000, France

Wufan Chen is with the School of Biomedical Engineering, Southern Medical University, Guangzhou, 510515, China

Corresponding Author : Limin Luo, e-mail: luo.list@seu.edu.cn

Nevertheless, this reconstruction becomes an ill-posed problem when it is carried out from a limited number of projections or has to deal with inconsistencies generated by the presence of noise in the measurements. Several applications are concerned with this problem such as low-dose computed tomography (CT), Positron Emission Tomography (PET) and Single Positron Emission Computerized Tomography (SPECT). The question of how to overcome such ill-posedness has been widely studied in the past twenty years [7-14]. Among the proposed solutions, a large number of work suggested solving it by imposing a priori knowledge as a regularization constraint in particular within a Bayesian reconstruction framework. The regularization prior, associated with the likelihood function, provide thus an approximate solution of the reconstruction problem as the maximum a priori (MAP) estimate that maximizes the log of the posterior probability [15].

A large variety of prior has been proposed in the literature. Quadratic priors such as Gaussian-Markov prior or the convex quadratic membrane (QM) prior [8,15], have the particularity to bring in a global smoothness of the solution with a tendency in blurring edges in the image. To preserve edges in regularization, some nonquadratic priors with nonquadratic energy potentials have been widely applied in tomographic reconstructions [10-14]. Some classic edge-preserving nonquadratic priors are the Huber potential function [8], Median Root Prior (MRP) [10] and TV prior [12]. The main drawbacks of these edge-preserving priors are the unguaranteed convexity, and the easily produced blocky artifacts when the measurements suffer from heavy noise [8,11-14].

In [7-15], prior computation usually relies on intensity differences between individual pixels and a simple spatial weighting strategy between neighboring pixels. Thus to take in more information on the image (geometry, statistics distribution of the structures of interest,), the idea has been to compute this difference between neighboring patches and apply weights that reflect the similarity between pixels within the neighborhood. This idea was inspired from Buades et al.'s work [16] on the nonlocal mean filter designed for image denoising. Our previous work [13] and [17] proposed thereby in the prior function, an adaptive weight as a decreasing function of the Euclidean distance between two square neighboring patches. This weight expressed a similarity between the neighboring pixels and was updated using the latest estimated image in a one step late (OSL) way. This method is however highly costly in computing all the patch distances and may provide inconsistent posterior energy function with some problem of convergence in some situations.

This paper aims at solving these problems by proposing a novel Patch Similarity based Mixture (***PSM***) prior model within a new convergent image/weight joint reconstruction algorithm. This PSM prior model provides as in [17], a weighted sum of the distances between patches surrounding each pair of neighboring pixels but weights are now estimated using a constrained entropy maximization, in which a normalization is introduced to preserve the balance between the regularization and image estimate. This proposed PSM prior and the statistical framework for MAP reconstruction are described in section II. To overcome the induced computational burden, a fast implementation of the most time consuming part of the algorithm is performed in Graphics programming Units (GPU) using Compute Unified Device Architecture (CUDA) framework [18-20]. An comparative evaluation of this algorithm was afterwards conducted both visually and quantitatively on simulated and real CT data. This evaluation included a study on the iteration monotonicity and sensitivity of the result according to the sizes of the comparing patches

and the searching neighborhood windows. The set of results are given in section III.

II. THEORY

*A. The Patch Similarity Based Mixture (**PSM**) Prior Model*

A MAP estimate of the image can be computed as the maximization of the posterior probability $P(f|g)$, which is defined by:

$$P(f|g) \propto P(g|f)\, P(f) \tag{1}$$

$$P(f) = Z^{-1} \times \exp(-\beta U(f)) \tag{2}$$

$$U(f) = \sum_j U_j(f) = \sum_j \sum_{b \in N_j} w_{bj} v(f_b - f_j) \tag{3}$$

where, f_j is the estimate for a pixel j in the image space, the potential function $v(f)$ relates to an intensity difference between the pixel j and the neighboring pixel b. N_j denotes the neighborhood window centered at pixel j. In traditional prior such as the Huber one, w_{bj} defines the weighting coefficient that is put to 1 for a neighboring pixel located in the horizontal and vertical direction and $1/\sqrt{2}$ when it refers to a neighboring pixel whose location is in the diagonal direction. They also usually express an intensity difference between individual pixels within local neighborhoods [8-15]. This expression is often inadequate to characterize large density distribution variations on the image. But this issue is made difficult by the fact we have to simultaneously satisfy a double constraint that is the image has to be locally smoothed while allowing sharp intensity transitions that correspond to tissue, organ or lesion boundaries. A way to take into account this double constraint is to introduce adaptive weights in the smoothing prior so that the potential function $v(f)$ may be set to 0 in case of dissimilarity between neighboring pixels to avoid smoothing between these pixels and preserve thus the transitions at the boundary of two regions.

We proposed in [17] a spatially adaptive nonlocal (NL) prior model in which $v(t) = t^2/2$ was used and w in (3) was determined by a similarity criterion between patches surrounding each pair of neighboring pixels within a neighborhood window of size N (Fig. 1). In [21], weights w and image f were alternatively updated within a one step late (OSL) algorithm. However, this joint updated reconstruction algorithm although efficient in practice in providing a spatially adaptive regularization, could lead to inconsistent posterior energy function with no guaranty of convergence within each iteration due to the non convexity of the potential function [8].

To overcome this problem, we propose here a novel PSM prior model, in which regularization takes effect though penalizing the Gaussian weighted Euclidean distance $||G_\alpha * n_b - G_\alpha * n_j||$ between the patches n_b and n_j surrounding each pair of pixel j and its neighboring pixel b in neighborhood N_j. Here, G_α is a two-dimensional Gaussian kernel of preset standard deviation α. All the weights w in each N are considered unknown variables with nonlinear dependence on the image f. This PSM prior can be considered as a high-level MRF prior that models the high-order patterns around each pixel in neighborhood N [15], [21-22]. For each pair of pixel j and b in neighborhood N_j, w_{bj} is considered as the similarity probability between the two patches n_b and n_j surrounding

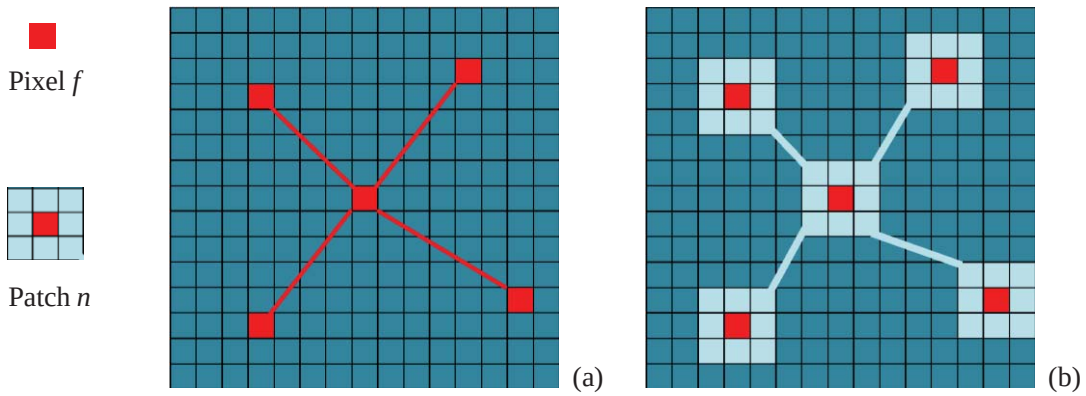


Fig. 1. (a),pixel-wise relations in pixel discontinuity (difference) calculation for traditional prior models; (b), Patch-wise relations in patch discontinuity (difference) calculation in the proposed PSM prior model.

pixels b and j [24]. We applied then the maximum entropy principle to derive a probability distribution which provides the best consistency in modeling the statistical properties of the image. As we looked for obtaining global patterns via stochastic propagation of local interactions, we choose thus the probability distribution that has the maximum entropy [23, 25]. This was done by introducing in the prior, an entropy term $\sum_j \sum_{b\in N_j} w_{bj} \ln w_{bj}$ on all the weights w_{bj} attached to each neighborhood window N_j with the following normalization constraint $\sum_{b\in N_j} w_{bj} = 1$. A large value of w_{bj} is associated with an elevated similarity probability between patches n_b and n_j and a low entropy prior energy and conversely, a small value of w_{bj} leads to a low similarity probability between patches n_b and n_j and a higher entropy prior energy.

B. Joint Weight/image Reconstruction Algorithm with the application in CT

The joint-MAP reconstruction algorithm can be therefore rewritten, according to the Bayesian theorem in order to build a posterior probability $P(f, w, \eta|g)$ that includes the parameters w and Lagrange-multiplier parameter η [15, 26] such as:

$$P(f, w, \eta|g) \propto P(g|f)P(f|w, \eta)P(w, \eta) \tag{4}$$

Substituting the log prior of the Gibbs distribution in (2) produces the objective posterior energy function:

$$\begin{aligned}\psi_\beta(f, w, \eta|g) &= \log P(g|f) + \log P(f|w, \eta) + \log P(w, \eta) = \; L(g, f) - \beta(\sum_j U_{PSM}(f_j, w, \eta|\hat{f})) \\ &= \; L(g, f) - \beta \sum_j ((\sum_{b\in N_j} w_{bj}||G_\alpha * n_b - G_\alpha * n_j|| + \lambda \sum_{b\in N_j} w_{bj} \ln w_{bj}) + \eta_j(\sum_{b\in N_j} w_{bj} - 1))\end{aligned} \tag{5}$$

where, $\hat{f}$ denotes the current estimated image. For each pixel j, this PSM prior uses the mixture prior energy $U_{PSM}(f_j, w, \eta|\hat{f})$ as a constrained combination of a pair-wise weighted patch similarity prior energy and an

entropy prior energy -$\sum_j \sum_{b\in N_j} w_{bj}\ln w_{bj}$. The parameter λ controls the entropy constraint on w , and is routinely set based on the noise level in measurement data.

The proposed PSM prior is applied in statistical CT image reconstruction. We model the scanned photon counts vectors as Poisson random variables. The likelihood function $P_L(g/f)$ for a given CT image f, is described by the independent Poisson distribution [4]:

$$P_L(g|f) = \prod_{i=1}^{I} \left[(\bar{g}_i\,(f)^{g_i}/g_i!)\exp\left(-\bar{g}_i\,(f)\right)\right] \tag{6}$$

with the mean

$$\bar{g}_i\,(f) = d_i \exp\left(-\left[Af\right]_i\right) = d_i \exp\left(-\sum_j^J a_{ij} f_j\right) \tag{7}$$

where, g_i is the measured photon counts detected by projection i can be modeled as Poisson random variables with $\bar{g}_i$ a function of the underlying attenuation map [1-4]. I is the number of projection lines, J is the number of image pixels and a_{ij} denotes the probability that a transmission from pixel j is detected by the detector i. d_i factors represents the photon number detected by detector i in the absence of the absorber. Thus the log-likelihood function for obtaining g from f is:

$$L(g,f) = \log P_L(g|f) = \sum_i \left(g_i \log \bar{g}_i(f) - \bar{g}_i(f) - \log g_i!\right) \tag{8}$$

Then, the objective CT image f and the weights w can be jointly learned through maximizing the $\psi_\beta(f,w,\eta|g)$ in (5) with respect to each element of f and w alternatively [26]:

1) Weight update (with fixed image $\hat{f}$): $\arg\max_w(-U_{PSM}(w,\eta|\hat{f}))$

Let $\hat{f}$ denotes the current image estimate, each weight w_{bj} is updated by solving $\arg\max_{w_{bj}}(-U_{PSM}(w,\eta|\hat{f}), b\in N_j)$:

$$w_{bj} = \arg\max_{w_{bj}}\left(-\sum_{b\in N_j} w_{bj}||G_\alpha * n_b - G_\alpha * n_j|| - \lambda\sum_{b\in N_j} w_{bj}\ln w_{bj} - \eta_j\left(\sum_{b\in N_j} w_{bj} - 1\right)\right) \tag{9}$$

w_{bj} is computed in the following way:

$$\left.\begin{array}{l} \frac{\partial U_{PSM}(\hat{f},w,\eta)}{\partial w_{bj}} = 0 \Rightarrow \|G_\alpha * n_b - G_\alpha * n_j\| + \lambda(1+\ln w_{bj}) + \eta_j = 0 \\ \frac{\partial U_{PSM}(\hat{f},w,\eta)}{\partial \eta_j} = 0 \Rightarrow \sum\limits_{b\in N_j} w_{bj} - 1 = 0 \end{array}\right\}$$

$$\Rightarrow \left\{\begin{array}{l} w_{bj} = \frac{\exp(-\|G_\alpha * n_b - G_\alpha * n_j\|/\lambda)}{\sum\limits_{b'\in N_j}\exp(-\|G_\alpha * n_{b'} - G_\alpha * n_j\|/\lambda)} \\ \eta_j = -\|G_\alpha * n_b - G_\alpha * n_j\| - \lambda - \lambda\ln\frac{\exp(-\|G_\alpha * n_b - G_\alpha * n_j\|/\lambda)}{\sum\limits_{b'\in N_j}\exp(-\|G_\alpha * n_{b'} - G_\alpha * n_j\|/\lambda)} \end{array}\right. \tag{10}$$

Here, $\|G_\alpha * n_b - G_\alpha * n_j\|$ can be easily calculated using current image estimate $\hat{f}$. From (10), we can see that each w_{bj} is determined not only from the similarity metrics between patches n_b and n_j, but also with the summed similarity metrics between all pairs of patches. So w_{bj} is different from w_{jb}.

2) Image update (with fixed weight $\hat{w}$): $\arg\max_f \psi_\beta(f, \hat{w}, \hat{\eta}|g)$

With the estimated weight denoted by $\hat{w}$ and the terms with no f omitted, the $\psi_\beta(f, \hat{w}|g)$ becomes

$$\psi_\beta(f, \hat{w}|g) = L(g, f) - \beta \sum_j \Big(\sum_{b \in N_j} \hat{w}_{bj} ||G_\alpha * n_b - G_\alpha * n_j|| \Big) \tag{11}$$

Based on the theory of pseudo-likelihood and Iterated Conditional Mode (ICM) in [15], we can get each f_j by solving the following factorized problem:

$$\arg\max_{f_j} \psi_\beta(f_j, \hat{w}_{bj}|g, \hat{f}) \Rightarrow \arg\max_{f_j} L\left(f, g\right) - \beta \Big(\sum_{b \in N_j} \hat{w}_{bj} ||G_\alpha * n_b - G_\alpha * n_j|| \Big) \tag{12}$$

The discretization property of the image pixels makes it suitable to apply the Iterative Coordinate Descent (ICD) algorithm to solve (12) [27-28]. So in this study we use the Paraboloidal surrogate coordinate ascent (PSCA) reconstruction in [9] to solve image estimate at each pixel f_j. As an ICD optimization procedure, the PSCA algorithm applies quadratic paraboloidal surrogate to achieve efficient reconstruction in transmission tomography [9]. This PSCA method has a close theoretical relation with the Optimization Transfer idea in [29], and the iterative scaling based CD optimization with a factorized resulting Newton step [28]. For the PSCA reconstruction, the quadratic paraboloidal surrogate of the log-likelihood term and the quadratic prior term allow exact Taylor expansions, which can lead to an efficient exact Newton update.

When the measured projection data g is also the observed image data f, this reconstruction problem degenerates into a image denoising problem and the above joint weight/image update algorithm leads to the iterative denoising algorithm applied in the Gaussian noise case [30]. It is trivial to prove that the $-U_{PSM}(w, \eta|\hat{f})$ in (16) is concave with respect to w. $\partial^2 L\left(g, f\right)/\partial f_b \partial f_j$ is known to negative definite [6,8]. Thus considering the convexity of $||G_\alpha * n_b - G_\alpha * n_j||$ in (11), and given the fixed computed weight $\hat{w}$, we conclude that the Hessian matrix of each $\psi_\beta(f, \hat{w}|g)$ in (11) is also negative definite. So each $\psi_\beta(f, w, \eta|g)$ is separately concave with respect to each f and w. Denoting $\hat{f}^k, \hat{w}^k, \hat{\eta}^k$ and $\hat{f}^{k+1}, \hat{w}^{k+1}, \hat{\eta}^{k+1}$, as the k^{th} and $k+1^{th}$ iterated estimations, the joint weight/image reconstruction algorithm yields to $\psi_\beta(\hat{f}^k, \hat{w}^{k+1}, \hat{\eta}^{k+1}|g) > \psi_\beta(\hat{f}^k, \hat{w}^k, \hat{\eta}^k|g)$ and $\psi_\beta(\hat{f}^{k+1}, \hat{w}^{k+1}, \hat{\eta}^{k+1}|g) > \psi_\beta(\hat{f}^k, \hat{w}^{k+1}, \hat{\eta}^{k+1}|g)$ ($\hat{\eta}$ is updated with $\hat{w}$ in (10)), respectively. We can therefore infer the following relation: $\psi_\beta(\hat{f}^{k+1}, \hat{w}^{k+1}, \hat{\eta}^{k+1}|g) > \psi_\beta(\hat{f}^k, \hat{w}^{k+1}, \hat{\eta}^{k+1}|g) > \psi_\beta(\hat{f}^k, \hat{w}^k, \hat{\eta}^k|g)$, which implies that convergence to local maximum is ensured for the iterated objective function $\psi_\beta(f, w, \eta|g)$.

Finally, the joint reconstruction algorithm is outlined as follows:

– – – – – – – – – – – – – – – – **Algorithm** – – – – – – – **Outline** – – – – – – – – – – – – – – – – – –

- **Define**: $L(g, f) = \sum_{i=1}^{I} h_i(l_i), \quad h_i\left(l_i\right) = \left(g_i \log\left(d_i \exp\left(-l_i\right)\right) - d_i \exp\left(-l_i\right)\ \right), \quad l_i = \left[Af\right]_i = \sum_j^J a_{ij} f_j$

  Initialize: set the initial image estimate $\hat{f}$, built projection matrix A based on system geometry, and set iteration number $niter$.

- ***For*** each iteration $k = 1, ..., niter$:

With current estimated image $\hat{f}$, update all $\hat{\eta}$ and $\hat{w}_{bj}$ using (10), and calculate $\hat{l}_i = \left[A\hat{f}\right]_i = \sum_j^J a_{ij}\hat{f}_j$.

For each pixel j in image f, based on the PSCA method in [9], we update the image estimate $\hat{f}_j$:

$$\dot{h}_i\left(\hat{l}_i\right) = \left(1 \; - \; \frac{g_i}{d_i \exp\left(-\hat{l}_i\right)}\right) d_i \exp\left(-\hat{l}_i\right), \quad \ddot{h}_i\left(\hat{l}_i\right) = \frac{d_i g_i \exp\left(-\hat{l}_i\right)}{\left(d_i \exp\left(-\hat{l}_i\right)\right)^2} \; - \; d_i \exp\left(-\hat{l}_i\right)$$

$$c_i\left(\hat{l}_i\right) = \begin{cases} \frac{2}{\hat{l}_i^2}\left[h_i\left(\hat{l}_i\right) - h_i\left(0\right) - \hat{l}_i \dot{h}_i\left(\hat{l}_i\right)\right], & \hat{l}_i > 0 \\ -\ddot{h}_i\left(\hat{l}_i\right) & \hat{l}_i = 0, \end{cases}$$

$$\dot{U}_{PSM}(f_j, \hat{w}, \hat{\eta}|\hat{f}) = \frac{\partial}{\partial f_j} U_{PSM}(f_j, \hat{w}, \hat{\eta}|\hat{f})\Big|_{f_j=\hat{f}_j} = -2\sum_{b\in N_j} \frac{\hat{w}_{bj}(\hat{f}_b - \hat{f}_j)}{||G_\alpha * n_b - G_\alpha * n_j||}$$

$$\ddot{U}_{PSM}(f_j, \hat{w}, \hat{\eta}|\hat{f}) = \frac{\partial^2}{\partial f_j^2} U_{PSM}(f_j, \hat{w}, \hat{\eta}|\hat{f})\Big|_{f_j=\hat{f}_j} = 2\sum_{b\in N_j} \frac{\hat{w}_{bj}}{||G_\alpha * n_b - G_\alpha * n_j||}$$

We can update the image estimate f_j:

$$\hat{f}_j = \left[\hat{f}_j + \frac{\sum_{i=1}^I a_{ij}\dot{h}_i(\hat{l}_i) - \beta\dot{U}_{PSM}(f_j, \hat{w}, \hat{\eta}|\hat{f})}{\sum_{i=1}^I a_{ij}^2 c_i(\hat{l}_i) + \beta\ddot{U}_{PSM}(f_j, \hat{w}, \hat{\eta}|\hat{f})}\right]_+$$

End

End

– – – – – – – – – – – – – – – – **Algorithm** – – – – – – – – **Outline** – – – – – – – – – – – – – – – – – –

C. Graphic Hardware Acceleration of the PSM Prior Computation

As illustrated in above algorithm outline, the PSM prior reconstruction involves pixelwise calculations of the $\dot{U}_{PSM}(f_j, \hat{w}, \hat{\eta}|\hat{f})$ and $\ddot{U}_{PSM}(f_j, \hat{w}, \hat{\eta}|\hat{f})$, in which intensive weight computations are required for each pair of patches with centers in each N_j over the whole image. This leads to a computational burden that is not compatible with the medical practice.

CUDA offers an unified hardware and software solution for parallel computing on NIVIA GPUs supporting the standard C/C++ programming language together with high performance computing numerical libraries [18]. A CUDA computation model is broken down into connected CPU (host) and GPU (device) parts. The direct heterogeneous CPU-GPU (Graphic Processing Unit) parallelism technique can be used to develop and deploy kernel functions for computations. The host executes code sequentially for GPU initialization, communication, and making kernel calls, and the GPU executes the kernel functions. Thus we can make well use of the parallel power to achieve time optimized result. Once the data loaded into the GPU memory, multi-kernel functions allow processing the data simultaneously [18-20]. The CPU calls kernel functions and each call results in the creation of threads and blocks in the GPU.

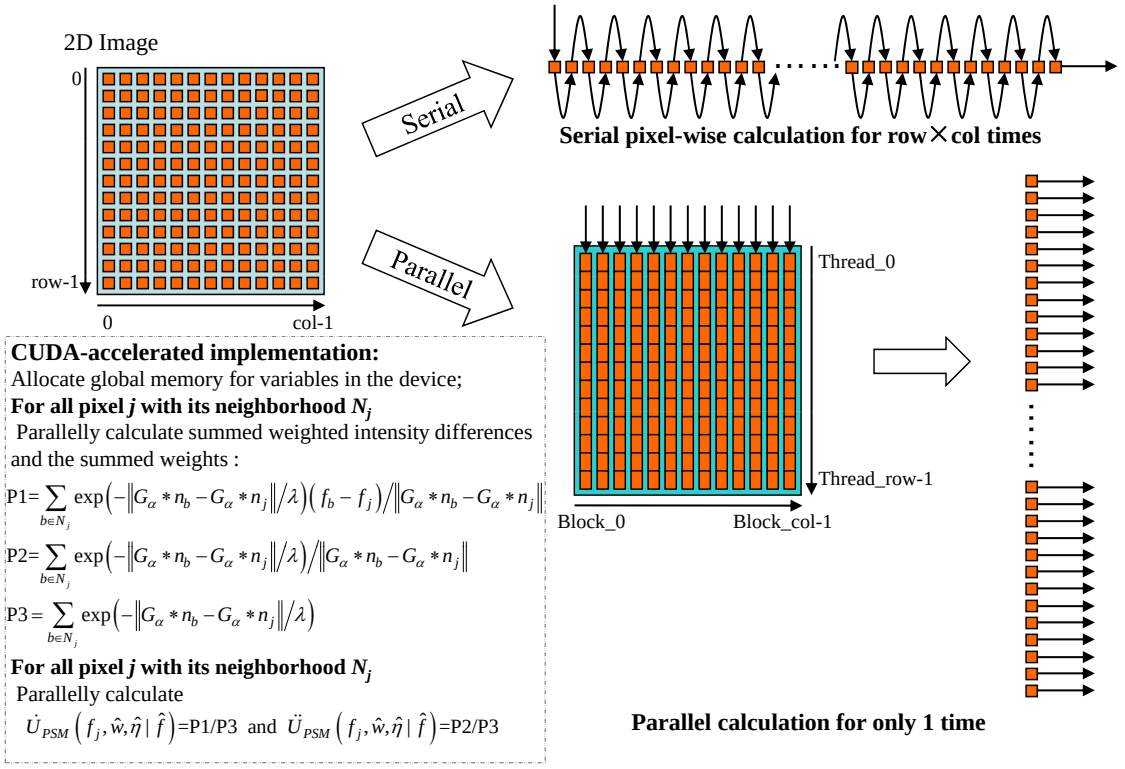


Fig. 2. CUDA acceleration of the computation of the $\dot{U}_{PSM}(f_j, \hat{w}, \hat{\eta}|\hat{f})$ and $\ddot{U}_{PSM}(f_j, \hat{w}, \hat{\eta}|\hat{f})$ in the PSM prior reconstruction: Serial vs parallel pixel-wise computation

We exploited therefore this environment to improve the performance of the proposed reconstruction, proposing a parallel implementing of the PSM prior. The structure of the CUDA implementation is illustrated in Fig.2. The total number of blocks in the grid is given by image row size and the total number of threads in each block by image column size. Therefore, at each iteration in the PSM prior reconstruction, all the threads in this block-grid structure are executed in parallel to perform all the pixelwise operations i.e those involved in the calculations of the weights for all patch pairs and then the pixelwise functions $\dot{U}_{PSM}(f_j, \hat{w}, \hat{\eta}|\hat{f})$ and $\ddot{U}_{PSM}(f_j, \hat{w}, \hat{\eta}|\hat{f})$.

In addition, we devised a more efficient method to calculate the distance dis_{bj} between the two comparing patches n_b and n_j in (11) in each iteration. In the following (13), we expand the weighted patch distance in (11) into **two summations** of the G_α weighted squared intensities in n_b and n_j, and **one summations** of the G_α weighted dot product between n_b and n_j:

$$\begin{aligned} ||G_\alpha * n_b - G_\alpha * n_j|| &= \sqrt{\sum_l (\alpha_l f_{l\in n_b} - \alpha_l f_{l\in n_j})^2} = \sqrt{\sum_l \alpha_l{}^2 (f^2_{l\in n_b} + f^2_{l\in n_j} - 2 f_{l\in n_b} \times f_{l\in n_j})} \\ &= \sqrt{\sum_l \alpha_l{}^2 f^2_{l\in n_b} + \sum_l \alpha_l{}^2 f^2_{l\in n_j} - 2\sum_l \alpha_l{}^2 (f_{l\in n_b} \times f_{l\in n_j})} \end{aligned} \tag{13}$$

Here, under the square root operator, the two first sums, that relates to each patch n_b and n_j, can be preliminary

computed before performing the pixel-wise calculation at each iteration. The third sum related to the intensity dot product between n_b and n_j can also be considered as one element of the convolution matrix between the central patch n_j and the other patches in the neighborhood N. Such convolution can be easily obtained with efficient multiplications, FFT (Fast Fourier Transform) and IFFT (Inverse FFT) operators. In addition, considering that the distance $||G_\alpha * n_b - G_\alpha * n_j||$ is equal to the distance $||G_\alpha * n_j - G_\alpha * n_b||$, we can save time by computing one distance on two.

III. EXPERIMENTATION AND ANALYSIS

In order to assess the algorithm effectiveness and speed, we present results on simulated and real data successively. We compare these results with two kinds of reconstruction methods, which are the conventional FBP and Bayesian-based algorithms respectively. For each of them, we considered different filters and priors. The list below gives the characteristics of each method:

1) FBP1 rec.: Filtered Backprojection reconstruction considering a ramp filter and a cutoff frequency equal to the Nyquist frequency;
2) FBP2 rec.: Filtered Backprojection reconstruction considering a hamming window and a cutoff frequency equal to 80% of the Nyquist frequency;
3) TV prior rec. : MAP reconstruction Algorithm based on a Paraboloidal Surrogates optimization procedure [9] and using a TV prior: $U(f) = \sum_j \sqrt{\left(\Delta_j^h f\right)^2 + \left(\Delta_j^v f\right)^2 + \epsilon^2}$;
4) Huber prior rec. : MAP reconstruction Algorithm based on a Paraboloidal Surrogates optimization procedure [9] and using a Huber prior : $U(f) = \sum_j U_j(f) = \sum_j \sum_{b \in N_j} w_{bj} v(f_b - f_j)$
   with $v(t) = \begin{cases} t^2/2, & |t| \leq \gamma \\ \gamma|t| - \gamma^2/2, & |t| > \gamma \end{cases}$ (γ is the bulit-in parameter controlling the sensitivity to noise);
5) MRP rec. : Bayesian-based Median Root Prior reconstruction algorithm [10];
6) PSM prior rec. : Above joint MAP algorithm applying the PSM prior with 7×7 patch n and 11×11 neighborhood N.

In all the reconstructions, the negative intensities in reconstructed images are set to 0 to preserve intensity positivity. The stopping criterion, for reconstructions using TV, Huber and MRP prior, is given by the condition (k denotes the iteration number):

$$\psi_\beta(f^{k+1}|g) - \psi_\beta(f^k|g) < 0.99(\psi_\beta(f^k|g) - \psi_\beta(f^{k-1}|g)) \tag{14}$$

In the case of the PSM prior based reconstruction, this criterion relies on the current estimated image intensities and weights (k denotes the iteration number):

$$\psi_\beta(f^{k+1}, w^{k+1}|g) - \psi_\beta(f^k, w^k|g) < 0.99(\psi_\beta(f^k, w^k|g) - \psi_\beta(f^{k-1}, w^{k-1}|g)) \tag{15}$$

In practice, between 800 and 1000 iterations are required for all theses methods to reach the stopping conditions. For the proposed PSM prior reconstruction, we use a 1-value unitary image as the initial condition because the streak

artifacts in FBP initial condition tends to mess up the whole iteration process. We will illustrate this in the later part of the paper. It is noted in [31] that, though showing a relative low convergence in the early iterations, reconstructions with a uniform initial condition are ultimately able to converge to results at least no worse than the ones with a FBP initial condition. For equal comparison, other Bayesian reconstructions also use 1-value unitary images as the initial conditions. Routinely, larger λ is set to suppress more noise when the noise level in the measured sinogram is higher, so we set λ as $\delta * var_s$ with sinogram variance var_s. It is hard to derive a deterministic relation between sinogram variances and the λ that will lead to the best reconstructions in terms of SNR or visual criterions. So we choose suitable λ by modulating the δ in $\delta * var_s$. In practical experiment, we found the suitable range for the parameter is from 8 to 14. This parameter determination strategy was also used in the Nonlocal means denoising in [16].

The algorithms have been written in hybird MATLAB and C++ language using the MATLAB release R2009b and NVIDIA CUDA libraries. The Operating System is Linux, the CPU core is INTEL E6600 and the GPU card is a NVIDIA GTX465 (Core Clock: 625MHz, Shader Clock: 1250MHz, Stream Processors: PCI Express 2.0×16 Memory: Effective Memory Clock: 3240MHz Memory Size: 1GB GDDR5 Memory). We build the projection matrix based on ASPIRE software system of Professor fessler's group was used as the programming platform [32].

A. Reconstruction Results on Simulated Data

Two synthetic phantoms have been considered that were both gray scale images whose intensity ranged within the [0-255] interval with a resolution of 8 bit 256×256 pixels and a pixel size of $1mm \times 1mm$. The first phantom was made of an ellipsoid of uniform background in which a cold square and circle was introduced, the circle having blurred edges and including a smaller hot square for simulating a lesion. The second phantom was a modified shepp-logan phantom in which 4 straight lines of different width was added. We modeled a monochromatic single slice detector CT scanner with a parallel beam geometry. We simulated a CT scan acquisitions with 367 radial bins and 360 angles uniformly spaced to produce a sinogram comprising 360 angular projections, each containing 367 linear samples. For the first phantom, we simulated a low-dose sinogram with a total count 0.7×10^8 (further named Phantom1-SL), and considered sinograms with two levels of total counts for the second phantom :1.2×10^8 and 4×10^8 (further named Phantom2-SL and Phantom1-SH). All the sinograms were generated by adding a 10% zero mean Gaussian noise (with a variance of 0.5) to take into account the electronic and quantum noises. The d_i factor in (7) were generated using pseudo-random log normal variates with a standard deviation of 0.3 to take into account the characteristics of the energy-integrating sensor in the X-Ray CT detector. The two original phantoms are displayed in Fig. 3.

Parameter setting related to the Bayesian based reconstructions are given Table I for each simulated phantoms Phantom1-SL, Phantom2-SL and Phantom2-SH. We used the following SNR equation to modulate parameter settings and evaluate the reconstruction qualities:

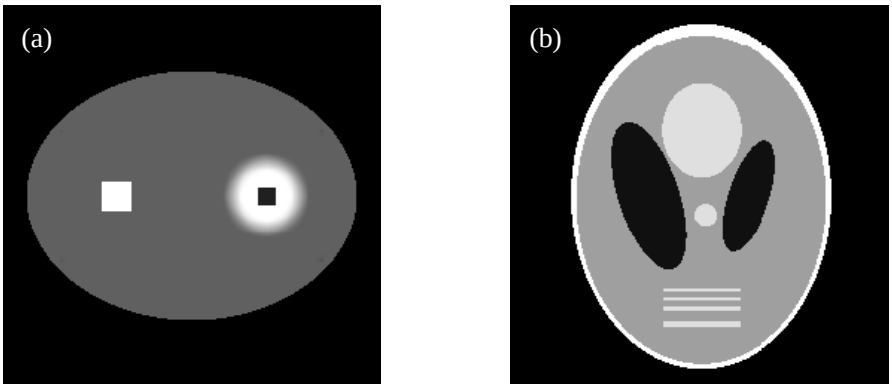


Fig. 3. Simulated phantom images.

Methods	Parameter setting		
	Phantom1-SL	**Phantom2-SL**	**Phantom2-SH**
TV prior reconstruction	$\beta = 4.2 \times 10^{-4}$, $\epsilon = 1e^{-10}$	$\beta = 5.4 \times 10^{-4}$, $\epsilon = 1 \times 10^{-10}$	$\beta = 3 \times 10^{-4}$, $\epsilon = 1 \times 10^{-10}$
Huber prior reconstruction	$\beta = 2.5 \times 10^{-2}$, $\gamma = 3$	$\beta = 1 \times 10^{-2}$, $\gamma = 1.6$	$\beta = 2.5 \times 10^{-3}$, $\gamma = 0.5$
MRP reconstruction	$\beta = 80$	$\beta = 40$	$\beta = 5$
PSM prior reconstruction	$\beta = 6.5 \times 10^{-2}$, $\lambda = 8.32$	$\beta = 8.5 \times 10^{-2}$, $\lambda = 7.62$	$\beta = 5 \times 10^{-2}$, $\lambda = 5.12$

TABLE I

PARAMETER SETTING IN THE EXPERIMENTS ON SIMULATED DATA

$$SNR = 10 \log_{10} \left(\frac{\sum_{j}^{J} \left(f_{phantom-j} - \bar{f}_{phantom} \right)^2}{\sum_{j}^{J} \left(f_{phantom-j} - f_j \right)^2} \right) \tag{16}$$

f_j, $f_{phantom-j}$ denote the intensity of the pixel j in the reconstructed image and the original phantom image respectively. $\bar{f}_{phantom}$ provides the mean of all the pixels in the original phantom image.

Fig. 4 shows the reconstruction results for the two sinograms simulated with a low dose scan (Phantom1-SL and Phantom2-SL). The visual comparison suggests that FBP1 method (with Ramp filter) suffers from severe noise effects and streak artifacts. Although FBP2 method (with Hanning filter) outperforms FBP1 method in noise suppression, many artifacts remains. Huber prior and MRP reconstructions tend to produce blocky piecewise smooth regions in the background and staircase effects around the edges. The straight lines introduced in the Shepp-logan phantom appear smoothed especially with the MRP method, the thinner lines being drowned out in the background. With TV prior, the results are improved, edges are better preserved even if we can observe a light staircase effect around the edges. Finally the best results are given with the PSM prior for which much less oversmoothing and staircase effects are observable.

Fig 5 provides the reconstruction results on the sinogram Phantom2-SL for which the sinogram has been simulated

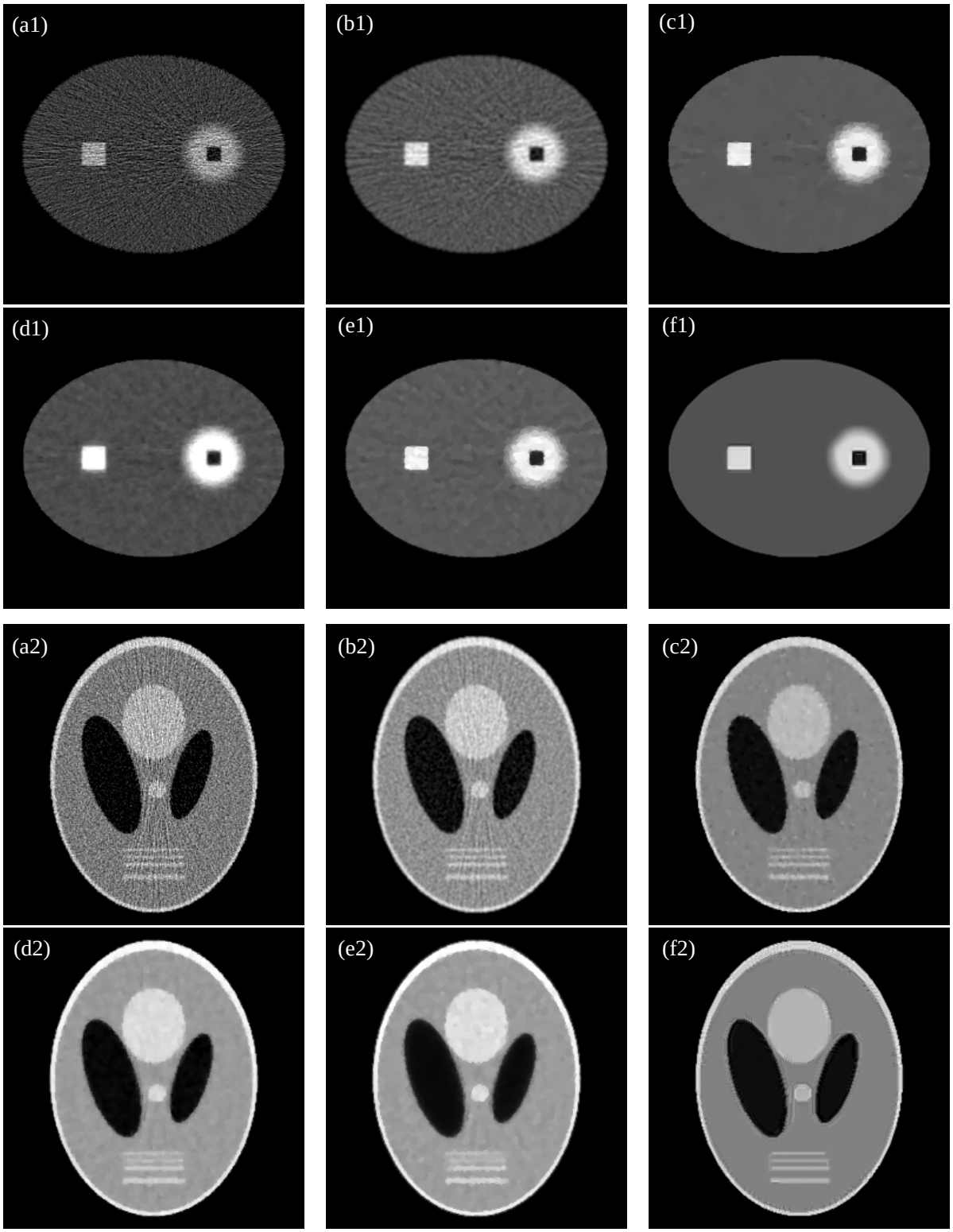


Fig. 4. Low dose reconstructions for the simulated data Phantom1-SL (a1-f1) and Phantom2-SL (a2-f2). (a1) and (a2), FBP1 reconstruction using Ramp filter; (b1) and (b2): FBP2 reconstruction using the Hanning filter with cutoff at 80% Nyquist frequency; (c1) and (c2): TV prior reconstruction; (d1) and (d2): Huber prior reconstruction; (e1) and (e2): MRP reconstruction; (f1) and (f2): The proposed PSM prior reconstructions.

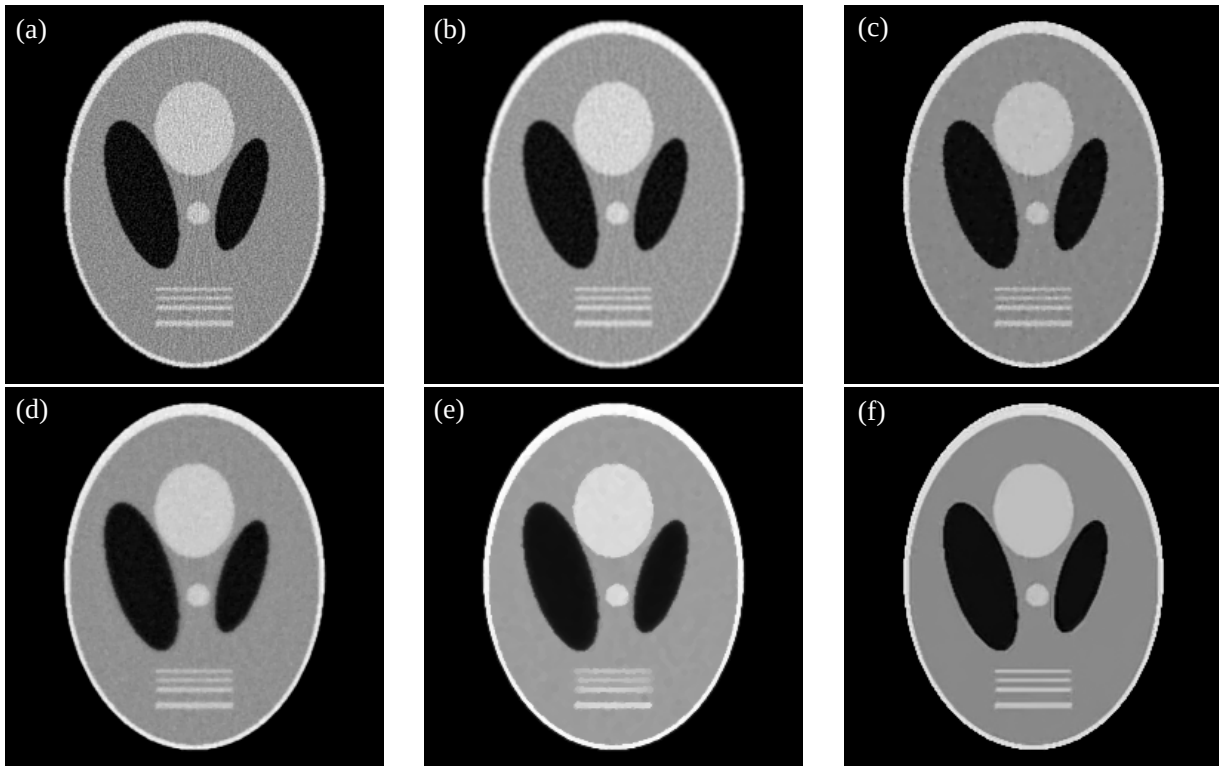


Fig. 5. High dose reconstructions for the simulated data Phantom2-SH. (a), FBP1 reconstruction using Ramp filter; (b), FBP2 reconstruction using the Hanning filter with cutoff at 80% Nyquist frequency;(c), TV prior reconstruction; (d), Huber prior reconstruction; (e), MRP reconstruction; (f), The proposed PSM prior reconstruction.

with a higher level of count. We can observe the same effects in a lower proportion since the magnitude of the noise is less important in the original measurements. FBP1 and FBP2 still produce some streak artifacts, Huber and MRP priors provide a smoothed image excepted that edges are better preserved in the image reconstructed with MRP prior, yielding thus a higher level contrast than in the former one. Smaller objects such as the straight lines appears nevertheless relatively smoothed. TV prior performs better in preserving edges and PSM prior still produce the best result both in term of smoothing and edge preservation. In TABLE II, we list the calculated SNR of all the reconstructions with the simulated data. Obviously, the SNR comparisons indicate that Bayesian reconstructions can produce images with higher SNR than analytic FBP reconstructions, and the reconstructed images using the proposed prior can obtain the highest SNR among all the reconstruction methods.

B. Reconstruction Results on Clinical Data

Real data came from a pelvis scan of a 56 year aged patient. This dataset was acquired on a monoenergetic Siemens Somatom volume zoom 1 detector scanner using the following protocol: 100 mA, 120 KV, pixel size:$1mm \times 1mm$, slice thickness: 1 mm, detector cell spacing : 1 mm, source to isocenter distance: 408 mm and source to detector distance: 816 mm. The sinogram consisted of 256 radial bins and 360 angular views. The acquired sinogram was

	FBP1 rec.	FBP2 rec.	TV prior rec.	Huber prior rec.	MRP rec.	PSM prior rec.
Phantom1-SL	14.21	15.48	23.27	21.64	21.47	23.75
Phantom2-SL	11.73	15.08	19.66	18.03	16.54	20.75
Phantom2-SH	16.12	16.85	22.25	20.83	20.56	23.43

TABLE II

SNR BETWEEN THE RECONSTRUCTED PHANTOM IMAGES (IN FIG. 4 AND FIG. 5) AND THEIR GROUND TRUTH PHANTOM IMAGES.

Methods	Parameter setting
TV prior rec.	$\beta = 5 \times 10^{-3}$, $\epsilon = 1 \times 10^{-10}$
Huber prior rec.	$\beta = 4 \times 10^{-1}$, $\gamma = 0.95$
MRP rec.	$\beta = 8$
PSM prior rec.	$\beta = 1.5 \times 10^{-2}$, $\lambda = 7.94$

TABLE III

PARAMETER SETTING IN THE EXPERIMENTS ON CLINICAL DATA

normalized, corrected for attenuation and scatter and rebinned to 2D using Fourier rebinning. The d_i factor in (7) was set to 1. As for the simulated data, we performed a set of experiments to select the optimal parameters for each method and asked a radiologist to evaluate the reconstruction quality. Listed in TABLE III, the parameters were thus chosen according to a visual inspection and based on the capability of the method to preserve the original anatomical structures.

Fig.6 illustrates the reconstructions on real clinical data. All the images are displayed using a window level and width respectively equal to 60 and 250 HU (Housfield units). FBP1 (with Ramp filter) produces similar effects to those met on the synthetic phantoms and FBP2 (with Hanning filter, Fig.6(b)) provides an improved image with lower streak artifacts. For the other methods, the differences locate in the details. We can thus observe in a lower proportion, blocky piecewise smooth regions in the background with undesirable staircase effects around the edges in the reconstructed images (Fig.6(c),(d)and (e)) from TV prior, Huber prior and MRP reconstructions. However, we can still notice blocky piecewise smooth regions with the TV prior, Huber prior and MRP reconstructions in Fig.6(c),(d)and (e). TV prior reconstruction provides a smoothing background with a loss of resolution on the small structures while preserving the sharp transitions in the images (see transitions between bones and soft tissues). Notably, in the proposed reconstruction of pelvis data in Fig.6(f), compared to other results, the bones (the brightest part of the image), scrotum, muscles(the regions around the bones) and fat tissue(the outer region beneath the skin surface) can be better identified with clearer detail preservation (the parts pointed by red arrows).

To complete this qualitative analysis, each method was evaluated in term of noise suppression, artifact suppression, contrast preservation and global image quality, using a five-point scale (1 = unacceptable, 2 = substandard, 3 = acceptable, 4 = above average, 5 = excellent). This evaluation was realized by three radiologists working in the department of radiology at the Nanjing Hospital in China (X.D.Y., with 15 years of experience, X.H.Y., with 8 years

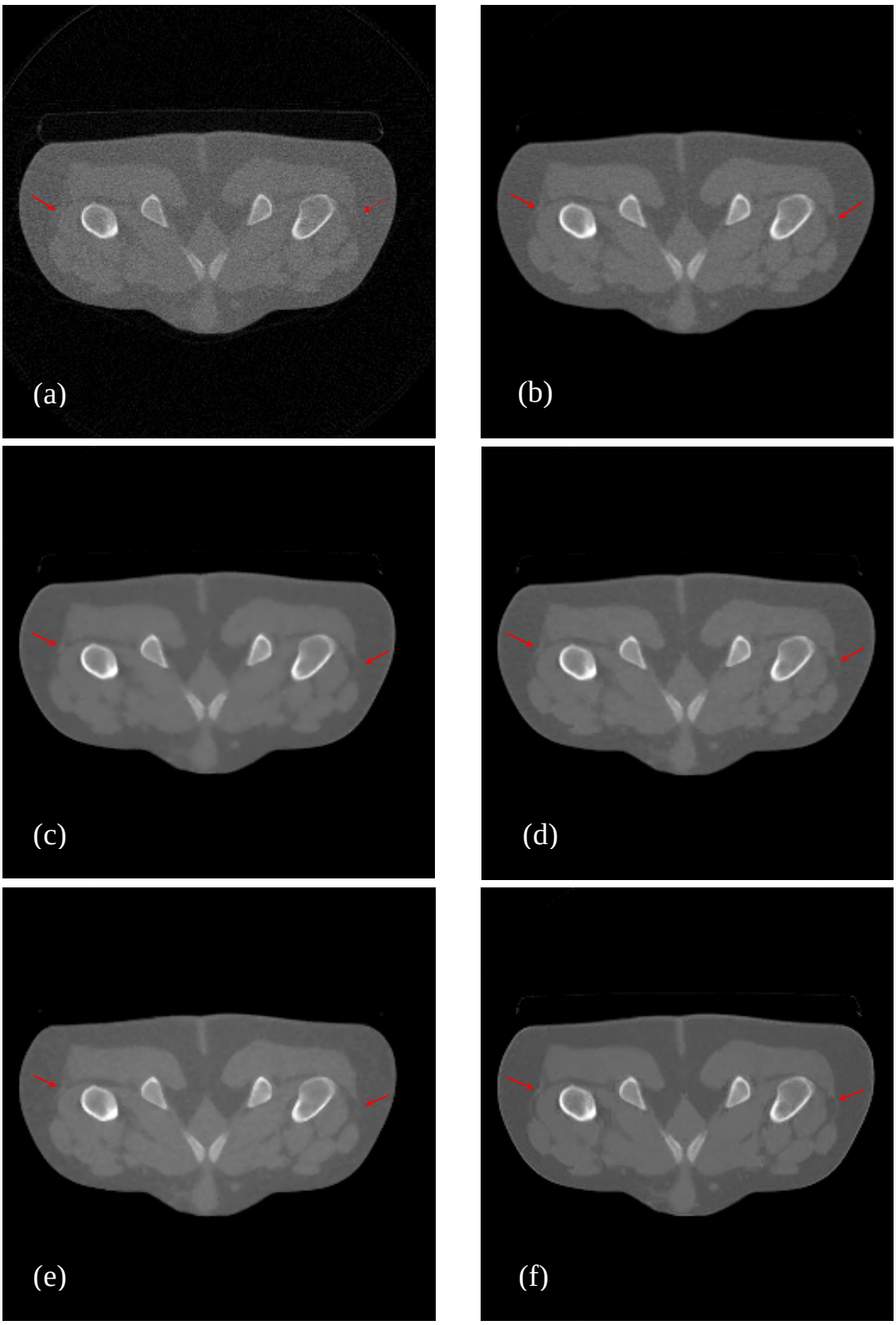


Fig. 6. FBP and Bayesian reconstructions using different priors for clinical pelvis data. (a), FBP1 reconstruction using Ramp filter; (b), FBP2 reconstruction using the Hanning filter with cutoff at 80% Nyquist frequency; (c), TV prior reconstruction; (d), Huber prior reconstruction; (e), MRP reconstruction; (f), the proposed PSM prior reconstruction.

	FBP1 rec.	FBP2 rec.	TV prior rec.	Huber prior rec.	MRP rec.	PSM prior rec.
Noise suppression	1.35	1.76	3.06	2.64	2.70	3.32
Artifact suppression	1.59	1.84	2.19	2.23	2.06	2.86
Contrast preservation	1.44	1.49	2.74	2.45	2.26	2.98
Global image quality	1.47	1.85	2.71	2.58	2.49	3.03
Global Mean Score	1.46	1.73	2.68	2.47	2.37	3.05

TABLE IV

QUALITY SCORES ASSOCIATED WITH THE RECONSTRUCTED CLINICAL IMAGES IN FIG. 6.

of experience and Y.M.D., with 5 years of experience). We asked them to independently assess the six randomized reconstructed images of the Figure 6. Un-stationary shapes that had a passive effects on image quality and included streak and blocky artifacts, were considered as artifacts. TABLE IV reports the quality scores of the different methods under the form of a mean score calculated on the responses of the three radiologists. The global mean score computed on the four criteria, for each method, hauls up the PSM method as the best one, followed by the the TV method. If we now look at each score, we can observe that the proposed PSM prior reconstruction is always superior to other reconstructions.

C. Computation Time Comparison

We measured then the acceleration factor, which is defined as the reduction factor in reconstruction time between the CUDA accelerated and the unaccelerated reconstruction algorithms. This study was carried out on all the algorithms and for the data Phantom1-SL using the same parameters as defined in above phantom experiment. The computation time is given in TABLE V for each algorithm in each situation: accelerated vs unaccelerated. For reference, the FBP reconstruction takes 21.89 CPU seconds. Without surprise, statistical methods appears much more expensive in computation time with a CPU time of about 4300 CPU seconds for the PSM prior method, and 900 CPU seconds for other prior methods. The GPU-based CUDA acceleration significantly reduces the overall computation time and allows a speed up factor of respectively 14 (for TV prior, Huber prior and MRP reconstructions) and 30 (for the proposed PSM reconstruction). The reason that the PSM prior reconstruction can obtain higher acceleration than other Bayesian reconstructions is that the prior involved calculation has a higher proportion in the PSM prior reconstruction. This parallelization was here limited to the weight calculation, and we would try to realize global acceleration of the algorithm (including forward/backward projections, dot multiplication between large matrixes [32-33])

D. Initial Conditions and Monotonicity for The Joint Image/Weight Reconstruction Algorithm

We carry out the PSM prior reconstruction on the Phantom1-SL data with the same parameters as previous experiment and a total number of iterations equal to 1000. Fig.7 (a) and (b) illustrate this situation with a 1-value unitary image and an image previously built from a FBP reconstruction with Ramp filter, respectively. Results show

	TV prior rec.	Huber prior rec.	MRP rec.	PSM prior rec.
CPU computation time for accelerated algorithms	66.54 sec	67.09 sec	55.74 sec	1474.42 sec
CPU computation time for Unaccelerated algorithms	909.62 sec	907.23 sec	835.57 sec	43090.83 sec
Acceleration factor	13	13	15	30

TABLE V

CPU TIMES (SECONDS) FOR EACH BAYESIAN BASED RECONSTRUCTION ALGORITHM

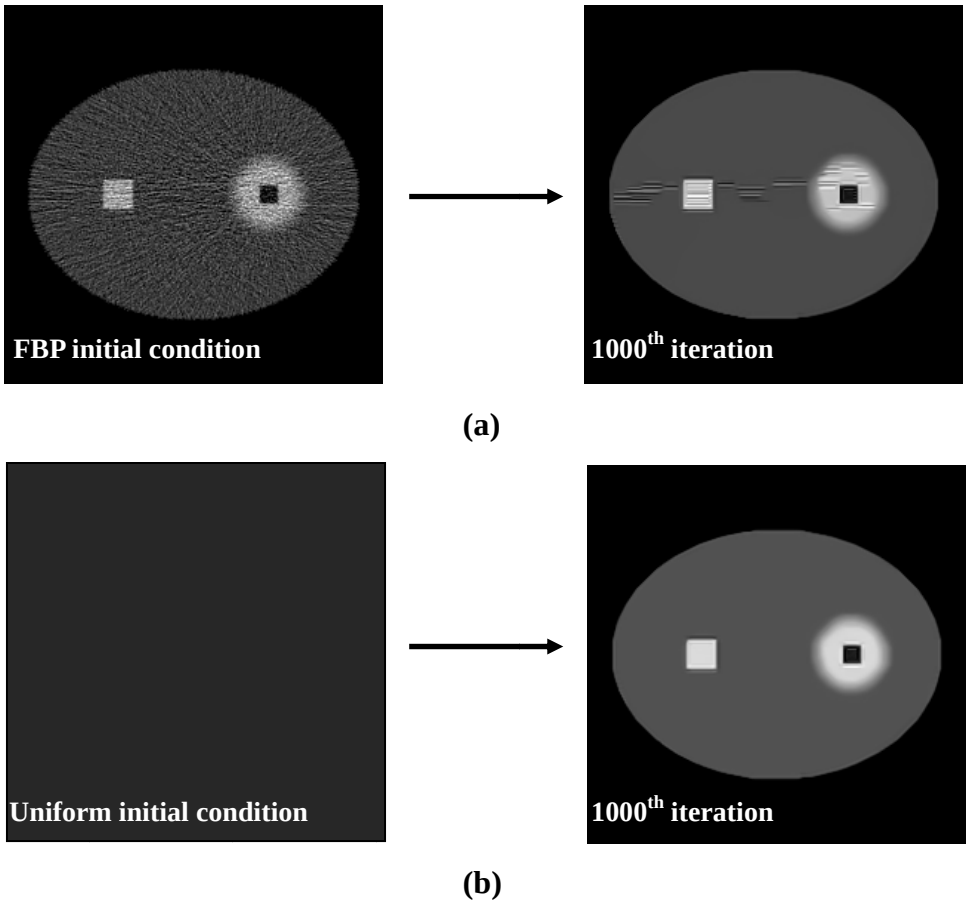


Fig. 7. Impact of different initial conditions on the reconstruction results. (a), FBP reconstructed initial image; (b), unitary initial image.

that using a FBP reconstructed initial image to initialize the reconstruction process, lead to the formation of streak artifacts in the final reconstructed image while a simple uniform initial image provides a good result.

We analyzed then the monotonicity of the PSM algorithm by calculating the posterior function energy $\psi_\beta(f, w, \eta|g)$ and the SNR with respect to the iteration numbers. Fig.8 depicts their evolutions according to the iteration numbers. It shows that we have a monotonic increase of the SNR and posterior function energy with the iteration progression. We can thus conclude that the convergence of the posterior energy is ensured in the iteration of the PSM algorithm.

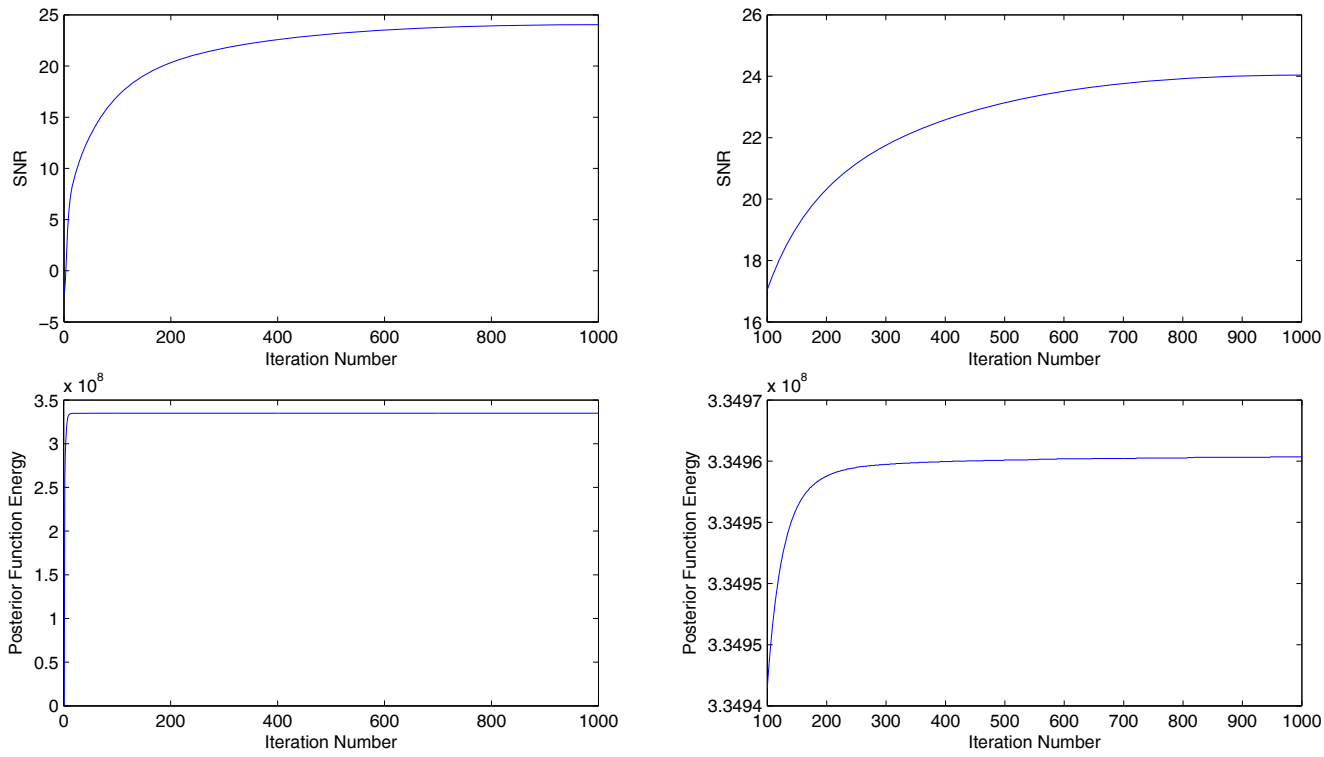


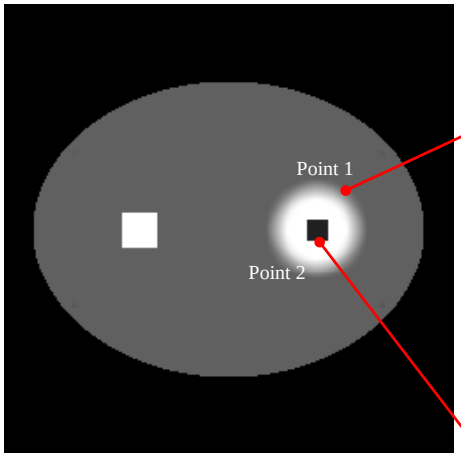
Fig. 8. Monotonicity Study : SNR (**upper left**:1-1000 iterations;**upper right**:101-1000 iterations) and posterior function energy (**lower left**:1-1000 iterations;**lower right**:101-1000 iterations) with respect to iteration number.

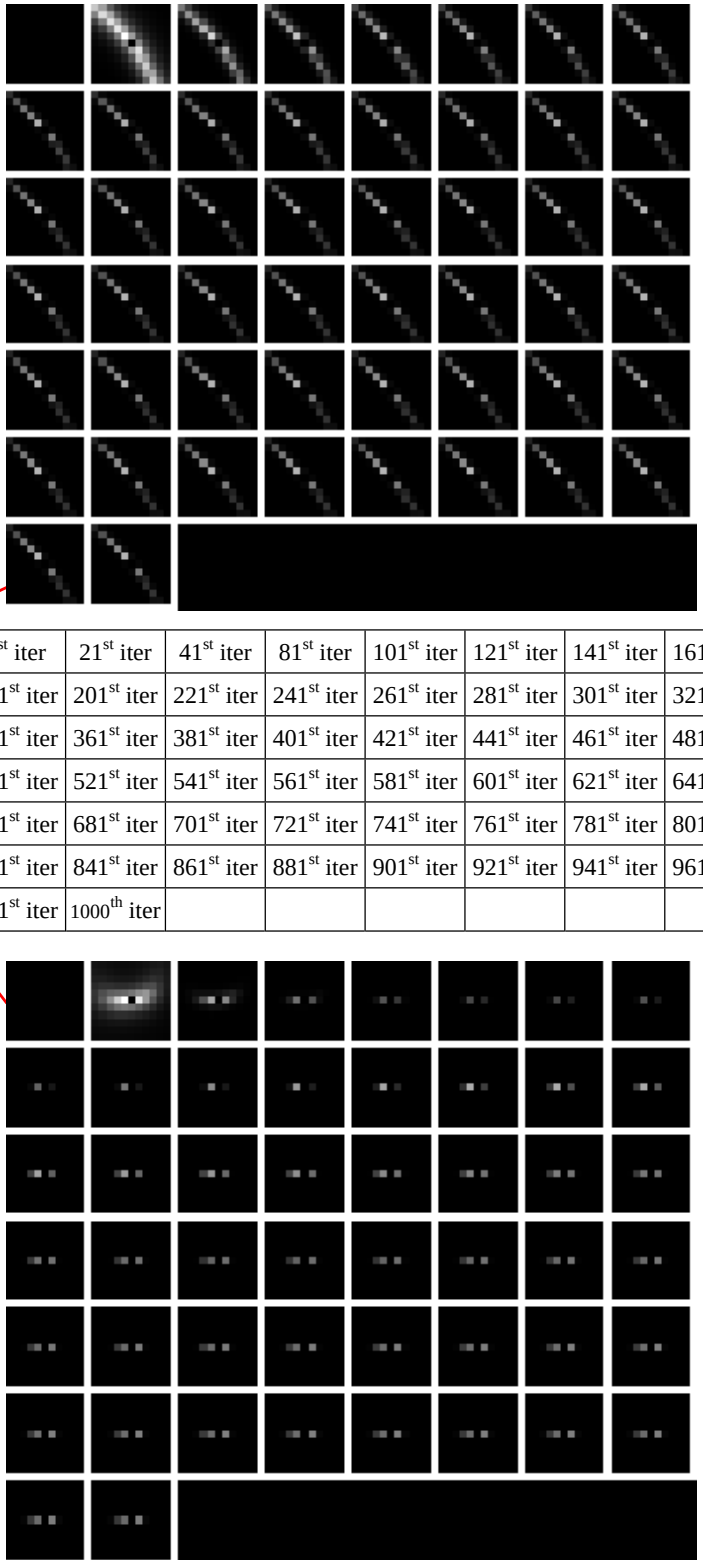
Fig. 9 provides the iterated weight maps for the two specified pixels in the image. It shows that the algorithm can achieve stable weight updates as iteration grows.

E. Study on Size Settings of N and n

We study here the impact of the size setting of N and n on the reconstruction qualities when using the proposed PSM prior. We performed the PSM reconstruction while making vary the sizes of N and n. We limited the size of n to 7×7 because a larger n might fail to present discriminative characteristics between the local structures and noise. The above CUDA accelerated algorithm was applied to perform reconstruction with data Phantom1-SL. Parameters β and λ were changed according to patch sizes based on the rule of SNR minimization. The best value for β was found to be 0.065, and λ was set to 8.32, 1.53 and 0.17 for n size respectively equals to 7×7, 3×3 and 1×1. Table VI provides for different couples (N,n), the calculated SNRs of the reconstructed images and the CPU times. When n is set to 1, the patch reduces to one single pixel and we recover the conventional pairewise TV-like priors. Fig. 10 illustrates the reconstructed images with zoomed regions of interests (in the red line boxes). We can see that when a small size of n is associated with a large size of N, the reconstruction provides blocky artifacts or staircase effects around the regions of interest. These effects decrease and the intensity transitions in the

\





1st iter	21st iter	41st iter	81st iter	101st iter	121st iter	141st iter	161st iter
181st iter	201st iter	221st iter	241st iter	261st iter	281st iter	301st iter	321st iter
341st iter	361st iter	381st iter	401st iter	421st iter	441st iter	461st iter	481st iter
501st iter	521st iter	541st iter	561st iter	581st iter	601st iter	621st iter	641st iter
661st iter	681st iter	701st iter	721st iter	741st iter	761st iter	781st iter	801st iter
821st iter	841st iter	861st iter	881st iter	901st iter	921st iter	941st iter	961st iter
981st iter	1000th iter						

Fig. 9. Phantom1-SL reconstruction using the PSM prior algorithm - The total number of iteration was set to 1000 : weight evolution map during iterations for two points (Point 1 and Point 2).

zoomed regions are better restored when the size of n increases. For this kind of images, the best SNR is given for the highest values of N and n. As expected, we can note in TABLE VI that more CPU time costs are needed for the reconstructions using the proposed prior with larger N and n. We observe in Fig. 10 and TABLE VI that no significantly distinct visual and SNR differences are made between 11×11 N and larger 13×13 N. This verifies our choice of 11×11 N and 7×7 n in this work.

N and n	SNR	CPU times	N and n	SNR	CPU times	N and n	SNR	CPU times
$(a), N_{3\times3}n_{1\times1}$	18.88	158.98 sec	$(b), N_{3\times3}n_{3\times3}$	19.62	378.59 sec	$(c), N_{3\times3}n_{7\times7}$	21.76	679.34 sec
$(d), N_{7\times7}n_{1\times1}$	18.95	119.12 sec	$(e), N_{7\times7}n_{3\times3}$	21.59	437.30 sec	$(f), N_{7\times7}n_{7\times7}$	22.78	542.09 sec
$(g), N_{11\times11}n_{1\times1}$	19.23	198.59 sec	$(h), N_{11\times11}n_{3\times3}$	21.05	632.42 sec	$(i), N_{11\times11}n_{7\times7}$	23.75	944.42 sec
$(j), N_{13\times13}n_{1\times1}$	19.19	491.54 sec	$(k), N_{13\times13}n_{3\times3}$	21.69	837.67 sec	$(l), N_{13\times13}n_{7\times7}$	24.06	1240.31 sec

TABLE VI

PHANTOM1-SL RECONSTRUCTION USING THE PSM PRIOR ALGORITHM: SNR AND CPU COMPUTATION TIME (SECONDS) ACCORDING TO THE SIZES OF THE PATCH n AND THE NEIGHBORHOOD N.

IV. CONCLUSIONS AND FUTURE WORK PLAN

We proposed in this paper a novel PSM prior which penalizes the distances between neighboring patches through a constrained mixture prior model. A convergent joint image/weight update algorithm, which estimates sequentially the image and the weights, has been also proposed to avoid the heuristically OSL weight determination. This algorithm has been applied in CT reconstructions with both synthetic phantom and real scan data. The quantitative and subjective analysis demonstrated the efficiency of the regularization in both suppressing noise and preserving detail as its supremacy over the classical edge-preserving TV prior, Huber prior and MRP algorithms. The uniform initial condition has shown to provide a better reconstruction result than the FBP initial condition and monotonic update of both image and weight has been demonstrated. The study on the sizes of the neighborhoods N and n has shown improved performances for larger values N and n.

The reconstruction using this new prior needs blockwise computations of the distances between patches over a large region, which leads to a computational burden. To overcome this problem, we suggested a graphic hardware acceleration to calculate the patch distance in a parallel way using the CUDA framework. A speed up of factor 30 has thus been achieved. This parallelization was here limited to the weight calculation. Next stage will be devoted to further accelerating the whole reconstruction by parallelizing the projection and image updates [33-34], developing more robust parameter estimations, and evaluating the proposed algorithm on more clinical CT data and other imaging modalities.

ACKNOWLEDGEMENT

This research was supported by National Basic Research Program of China under grant (2010CB732503),National Natural Science Foundation under grant (81000636), and the Project supported by Natural Science Foundation of

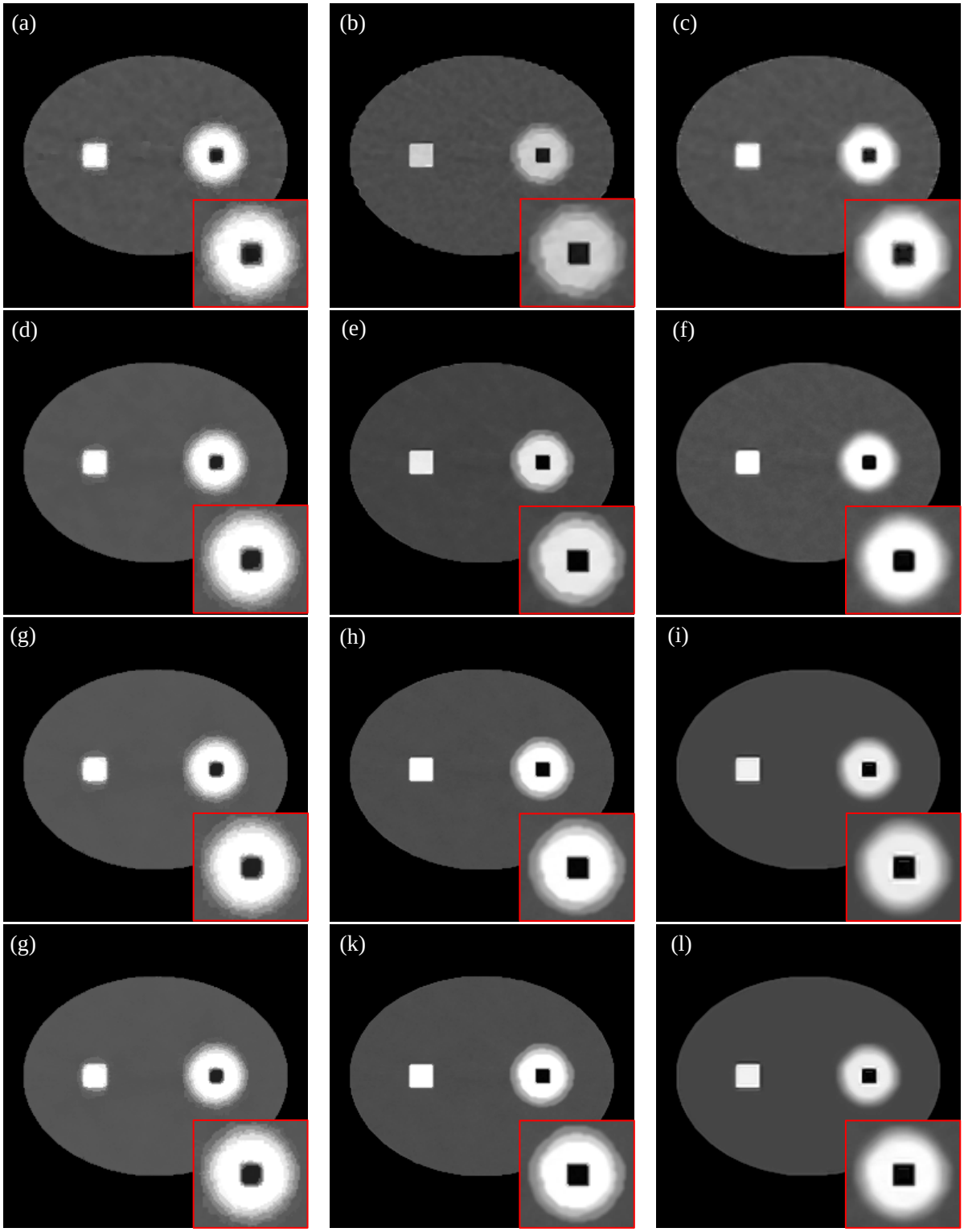


Fig. 10. Reconstruction results according to the sizes of N and n. Zoomed regions of interest are also illustrated in red line boxes. Each image is associated with the couple (N,n) having the same index in TABLE VI. Let's note that it is not easy to distinguish the differences between the images reconstructed with a size of N equal to 11×11 and 13×13 especially when n is equal to 7×7.

Jiangsu Province (BK2009012).

REFERENCES

[1] A.C Kat and M. Staney, "Principles of Computerized Tomographic imaging," New York: IEEE Press,177-201, 1988.

[2] C. Zhang, C. Li, L.V. Wang, "Fast and robust deconvolution-based image reconstruction for photoacoustic tomography in circular geometry: experimental validation," *IEEE Photonics Journal*, 2, pp. 57-66, 2010.

[3] F. J. Beekman and C. Kamphuis, "Ordered subset reconstruction for x-ray CT", *Phys. Med. Biol*, 46, pp. 1835-1855, 2001.

[4] I. A. Elbakri and J. A. Fessler, "Efficient and accurate likelihood for iterative image reconstruction in X-ray computed tomography," In Proc. SPIE 5032, Medical Imaging 2003: pp. 1839-50, 2003.

[5] L. A. Shepp and Y. Vardi, "Maximum likehood reconstruction for emission tomography", *IEEE Trans. on Medical Imaging*, vol. 1, 113-121,1982.

[6] K. Lange and J. A. Fessler, "Globally convergent algorithms for maximum a posteriori transmission tomography," *IEEE Trans. on Image Processing*, 4, pp. 1430-1438, 1995.

[7] M. Bertero, T. Poggio, and V. Torre, "Ill posed problems in early vision," *Proc. IEEE*, 76, pp. 869-889, 1988.

[8] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs smoothness," *IEEE Trans. on Medical Imaging*, 9, pp. 439-446, 1990.

[9] H. Erdoğan and J. A. Fessler, "Monotonic algorithms for transmission tomography," *IEEE Trans. on Medical Imaging*, 18, pp. 801-814,1999.

[10] W Chlewicki, F Hermansen and S B Hansen, "Noise reduction and convergence of Bayesian algorithms with blobs based on the Huber function and median root prior," *Phys. Med. Biol*, 49, pp. 4717-4730, 2004.

[11] C. A Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. on Image Processing*, vol. 2, pp. 296-310, 1993.

[12] V. Duval, J. F. Aujol, and Y. Gousseau, "The TVL1 model: A geometric point of view," *Multiscale Modeling and Simulation*, 8(1):154-189, 2009.

[13] Y. Chen, D. Gao, C. Nie, L. Luo, W. Chen, X. Yin, Y. Lin "Bayesian statistical reconstruction for low-dose X-ray computed tomography using an adaptive-weighting nonlocal prior," *Computerized Medical Imaging and Graph*, vol. 33, pp.495-500,2009.

[14] D. F. Yu and J. A. Fessler, "Edge-Preserving tomographic reconstruction with nonlocal regularization," *IEEE Trans. on Medical Imaging*, vol.21, pp. 159-173, 2002.

[15] Stan Z. Li, Markov Random Field Modeling in image Analysis, *Springer-Verlag*, Tokyo, 2001.

[16] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling and Simulation*, 4, pp. 490-530, 2005.

[17] Y. Chen, J. H. Ma, Q. Feng, L.M. Luo, P. C. Shi, and W. F. Chen, "Nonlocal prior Bayesian tomographic reconstruction," *Journal of Mathematical Imaging and Vision*, 30, pp.133-146, 2008.

[18] NVIDIA CUDA^{TM} Programming Guide(Version 3.0), *http://developer.download.nvidia.com/compute/cuda/3.0/toolkit/docs/NVIDIA CUDA Programming Guide.pdf*

[19] Accelerating MATLAB with CUDA Using MEX Files (White Paper), *http://developer.nvidia.com/object/matlab cuda.html*

[20] GPU Acceleration in MATLAB, *http://arch.eece.maine.edu/superme/images/8/8c/Final report.pdf*

[21] S.W. Lyu, "An implicit Markov random field model for the multi-scale oriented representations of natural images," IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 1919-1925.

[22] P. Winkler. Image Analysis, Random Fields And Markov Chain Monte Carlo Methods. Springer, 2nd edition, 2003.

[23] G. Peyr, S. Bougleux, L. Cohen, "Non-local regularization of inverse problems," European Conference on Computer Vision 2008, pp. 57-68.

[24] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," *Multiscale Modeling and Simulation*, 7(3), pp. 1005-1028, 2008.

[25] S. C. Zhu, Y. N. Wu and D.B. Mumford,"FRAME: filters, random field And Maximum Entropy:Towards a Unified Theory for Texture Modeling," *International Journal of Computer Vision*, 27(2) pp. 1-20, 1998

[26] I.T. Hsiao, A. Rangarajan, and G.R. Gindi, "Bayesian image reconstruction for transmission tomography using deterministic annealing," *Journal of Electronic Imaging*, 12(1), pp. 7-16, 2003.

[27] C. A. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Trans. on Image Processing*, 5(3), pp. 480-492, 1996.

[28] F. L. Huang, C. J Hsieh, K. W. Chang and C. J. Lin, "Iterative scaling and coordinate descent methods for maximum entropy models," *Journal of Machine Learning Research*, 11, pp. 815-848, 2010.

[29] K. Lange, D. R. Hunter, and I. Yang. "Optimization transfer using surrogate objective functions," *Journal of Computational and Graphical Statistics*, 9(1), pp. 1-20,2000.

[30] C. A. Deledalle, L. Denis and F. Tupin, "Iterative weighted maximum likelihood denoising with Probabilistic Patch-Based Weights," *IEEE Trans. on Image Processing*, 18(12), pp. 2661-2672, 2009.

[31] W. Zbijewski and F.J. Beekman, "Suppression of intensity transition artifacts in statistical X-ray computer tomography through Radon inversion initialization,"*Medical Physics*, 31, pp. 62-69, 2004.

[32] J. A. Fessler, "Aspire 3.0 user's guide: A sparse reconstruction library," Communication & Signal Processing Laboratory Technical Report No. 293, Department of Electrical and Computer Engineering, University of Michigan, Ann Arbor, 1998.

[33] H.Scherl,B. Keck, M. Kowarschik, et al,"Fast GPU-based CT reconstruction using the common unified device architecture (CUDA)," In: Proceedings of 2007 IEEE Nuclear Science Symposium Conference. Honolulu, Hawaii, 2007: 4464-4466.

[34] M. Knaup, S.Steckmann, M. Kachelrie, "GPU-based parallel-beam and cone-beam forward-and backprojection using CUDA," In: Proceedings of 2008 IEEE Nuclear Science Symposium Conference. Dresden, Germany, 2008: 5153-5157.